ABSTRACT

A process for modeling numerical data from a data set including collecting data for development of a model with a data acquisition module, processing the data to enhance its exploitability in a data preparation module, constructing a model by learning the processed data in a modeling module, evaluating the fit and robustness of the obtained model in a performance analysis module, adjusting the model parameters to select the optimal model in an optimization module, wherein the model is generated in the form of a D^{th} order polynomial of the variables used in input of the modeling module, by controlling the trade-off between the learning accuracy and the learning stability with the addition to the covariance matrix of a perturbation during calculation of the model in the form of the product of a scalar λ times a matrix H or in the form of a matrix H dependent on a vector of k parameters $\Lambda = (\lambda_1, \lambda_2, \ldots \lambda_k)$ where the order D of the polynomial and the scalar λ, or the vector of parameters Λ, are determined automatically during model adjustment by the optimization module by integrating an additional data partition step performed by a partition module which consists in constructing two preferably disjoint subsets: a first subset comprising training data used as a learning base for the modeling module and a second subset comprising generalization data destined to adjust the value of these parameters according to a model validity criterion obtained on data that did not participate in the training, and where the matrix H is a positive defined matrix of dimensions equal to the number p of input variables into the modeling module, plus one.